Logic in Visual Brain: Compute to Recognize Similarities

Formalized Anatomical and Neurophysiological Bases of Cognition

Andrzej W. Przybyszewski[1,2]

[1]Department of Neurology, UMass Medical School, 65 Lake Av. Worcester MA 01655, USA

[2]Polish Japanese Institute of Information Technology, Koszykowa 86, Warsaw 02008, Poland

[1]Andrzej.Przybyszewski@umassmed.edu; [2]Andrzej.Przybyszewski@pjwstk.edu.pl

***Abstract*-Object recognition is a complex neuronal process determined by interactions between several visual areas: from the retina, thalamus to the ventral visual pathway. These structures transform variable, single pixel signal in photoreceptors to a stable object representation in higher areas of the visual cortex. Neurons in macaque monkey area V4, midway in ventral stream, represent such stable shape detector. Traditionally these processes are described as feed forward hierarchy of increasing in size and complexity receptive fields (static spatiotemporal filters). A fundamental question in visual neuroscience is how these processes might identify an object or its elements in order to recognize it in new, unseen conditions? We propose a new approach to this problem by extending the classical definition of the receptive field (RF) to a fuzzy detector. Our RF modification is a consequence of the computational properties of the bottom-up and top-down pathways comparing stimulus with predictions. The "driver-type" logic (DTL) of bottom-up computations looks for large number of possible object parts (hypotheses), as object's elements are similar to RF properties. The optimal combination is chosen, in unsupervised, parallel, multi-hierarchical pathways by the "modulator-type" logic (MTL) of top-down computations. The DTL is related to anatomic divergence of ascending pathways and represents a large assembly of possible combinations of elements. Anatomical convergences of descending pathways determine selective property of the MTL. Such interaction between DTL (hypotheses) and MTL (predictions) gives the visual system universality of processing (with a high resolution of lower areas) vast number of possible visual cues and flexibility to choose right one (in agreement with an individual experience).**

Keywords-Fuzzy Detector; Ascending Descending Pathways; Object Categorization; Predictive Coding; Bayesian Cortical Computation; Rough Set Theory; Inconsistent Rules

I. INTRODUCTION

How slow and noisy brain's computations make our recognition so effective that it outperforms many times faster artificial intelligent (AI) systems? Can we at least find out what differences are in computations between these systems?

In this paper we try respond to above questions mainly in relationship to the electrophysiological data. On their basis we propose a model that has some analogies to predictive coding models: the Helmholtz machine [12], Rao and Ballard's predictive code model [51], and hierarchical Bayesian inference model [29]. However, in the Helmholtz machine model [12] the feedback mechanism was limited by the learning phase, and [29] was primary the linear model assuming that feedback serves to suppress activity in the early visual areas and only error residuals are projected to the higher areas. Therefore these models are not in agreement with our and others' experimental data [3, 7, 8, 43, 60].

In our everyday life we actively perceive only a small part of our environment. This part depends on our interest, which determines where we direct our eyes. This paper describes neurological mechanisms that determine how different brain structures may anticipate where are we going to look next in order to fulfill our needs or interest. There are two anatomically different pathways that interact in order to focus our attention on a specific object. One pathway has specific sub-cortical input (core cells) whereas another pathway has diffused sub-cortical inputs (matrix cells). The first pathway classifies objects on the basis of their visual attributes in contrast to the second pathway where classifications may be related to different motor activities like anticipation of eye movement or possibility to grasp an object, anticipation of obtaining higher value food reward, avoidance of the danger or obtaining of the pleasure. As the precision and meanings of these two pathways are different, in order to classify objects, we use a general model based on the rough set theory [34]. Our model by using similarities between objects and RF attributes takes into account differences between different anatomical pathways. On the basis of the similarity relation definition [39], we propose to classify objects by assembly of RF related granules that differentiate our method from that used in most AI applications.

It is generally accepted that responses in higher visual cortices can be significantly modulated by attention related to specific location, but it is not clear what the role of the higher visual areas in object recognition is.

A popular understanding on how neurological processes in the visual system lead to objects classification is based on generalization of simple and complex cell properties from visual area V1 as described by Hubel and Wiesel [23]. They proposed that an array of spatially aligned receptive fields (RFs) of LGN cells might give orientation sensitivity to V1 simple cell (SC), and that several phase (or position) shifted SCs with similar orientations convergent on the complex cell (CC). Such convergence might give spatial invariance in complex cells. On the basis of simple and complex cells properties, Fukushima [17] made simulation of a self-organizing network: cognition, and later introduced improved model with a position invariant property [18]. Networks with similar principles are still used nowadays in most models of the visual system. There are based on a first-order description of primary visual cortex V1 that consist of a collection of locally-normalized, threshold Gabor wavelet functions spanning a range of orientations and spatial frequencies [9, 28, 33]. More complex cells' properties arise in such linear models as summation of simple/complex cells from V1. There are many dissertations using this approach in different visual

areas: V1 till IT.

The linear combination of simple and complex cell RF attributes from areas V1, V2 may explain selectivity and position invariance properties of cells in area V4 [9, 52]. The main assumptions of above models are that simple units in higher areas (V4) generate selectivity for complex features or shapes by summation of units selective to different orientations and different receptive field sizes. Such linear, feed forward models can simulate certain sensitivity of V4 cells to complex object but cannot explain universality of higher brain areas to recognize complex objects in unseen conditions. Another problem with these models is that they do not take into account nonlinear properties of the complex cells such as, for example, overlapping of the on and off subfields [19, 26, 58]. Also some basic experimental findings of cell properties in area V4 like nonlinear interactions between subfields are not taken into account in above models [37].

II. METHODS

Below, we give formal definitions of an information system, similarities, object's attributes, and decision rules. An information system is a set of objects with their attributes put in a table. On this basis, one can find rules describing relationships between objects and their attributes. By quantifying attributes we can find similarities between objects assuming that attributes have certain ranges like for example orientation and orientation bandwidth.

A. Definition of an Information System

Rough set-based data analysis starts from a data table, called an information system. The information system contains data about objects of interest characterized in terms of some attributes. Often we distinguish in the information system condition and decision attributes. Such information system is called a decision table. The decision table describes decisions in terms of conditions that must be satisfied in order to carry out the decision specified in the decision table. With every decision table a set of decision rules, called a decision algorithm, can be associated. It is shown that every decision algorithm reveals some well-known probabilistic properties; in particular it satisfies the total probability theorem and Bayesian's theorem. These properties give a new method of drawing conclusions from data, without referring to prior and posterior probabilities, inherently associated with Bayesian reasoning.

After Pawlak [34], we define an information system as $S = (U, A)$, where U is a set of objects and A is set of attributes. In agreement with the Leibniz' principle we assume that objects are completely determined by their set of properties. If $a \in A$ and $u \in U$, the value $a(u)$ is a unique element of V (a value set). The indiscernibility relation of any subset B of A, or $IND(B)$, is defined as the equivalence relation whose elements $(u, w) \in IND(B)$ if $a(u) = a(w)$ for each $a \in B$, and $[u]_B$ - the equivalence class of u form *B-elementary granule*. A lower approximation $\underline{B}X$ of set $X \subseteq U$ is defined as $\underline{B}X = \{u \in U: [u]_B \subseteq X\}$. An upper approximation of X is defined as $\overline{B}X = \{u \in U: [u]_B \cap X \neq \varphi\}$. The set $BN_B(X) = \overline{B}X - \underline{B}X$ will be referred to as the *B-boundary region* of X. If the boundary region of X is the empty set then X is exact (crisp) with respect to B; otherwise if $BN_B(X) \neq \phi$ X is not exact (rough) with respect to B.

In this paper the universe U is a set of simple visual patterns that were used in our experiments [37, 43], which can be divided into equivalent indiscernibility classes related to their physically measured, computer generated attributes or B-elementary granules, where $B \in A$. The purpose of our research is to find how these objects are classified in the brain. Therefore, after Pawlak [34], we will modify the definition of the information system as $S = (U, C, D)$ where C and D are condition and decision attributes. Decision attributes will classify elementary granules in agreement with neurological responses from the specific visual brain area.

B. Definition of Similarity

In order to measure similarities between different objects (stimuli) quantitatively, we will introduce the rough inclusion measure [39].

A rough inclusion m is ternary relation, a subset of the product $U \times U \times [0, 1]$: $m(x, y, r)$, where x, y are individual objects, $r \in [0,1]$, satisfies following requirements:

1. $m(x, y, 1) \Leftrightarrow x \; ing \; y$
2. $m(x, y, 1) => [m(z, x, r) => m(z, y, r)];$
3. $m(x, y, r) \wedge s < r => m(x, y, s).$

The first condition means that m is extension of a notion of an element, which as equivalent to a part or subset in mereology is used as the notion of ingredient ing : $x \; ing \; y \Leftrightarrow y \vee x = y$; where a relation of being a part p satisfies the following conditions: 1. x p x there is no such x; 2. $x \; p \; y \wedge y \; p \; z => x \; p \; z$, which means that there is no such element as being part of itself, and it satisfies transitive property; the second condition is related to monotonicity of m and the third condition can be ready as "to degree at least r".

The family $\{m(x, y, r): r \in [0,1]\}$ can be seen as tolerance or similarity relation (it is even more general relation as a quasi-similarity class – see [39]). We can think about granule $g_m(x, r)$ as a list of objects y for which $m(x, y, r)$ holds. We notice that the rough inclusion is an extension of indiscernibility relation when defined as: $m(x, y, r)$ if $|a \in B: a(x) = a(y)\}|/|B| >= r$. That means that the number of attributes with the same value to all attributes is larger than r.

C. Object's Attributes

We will represent experimental data ([37]) in the following table. In the first column are neural measurements. Neurons are identified using numbers related to a collection of figures in the previous paper [37]. Different measurements of the same cell are denoted by additional letters (a, b …). For example, 11a denotes the first measurement of a neuron numbered 1 Fig. 1 of [37], 11b the second measurement, etc. Stimuli typically used in neuroscience have the following properties:

1. Orientation in degrees appears in the column labeled *o*, and orientation bandwidth is labeled by *ob*.

2. Spatial frequency is denoted as *sf* and spatial frequency bandwidth is *sfb*.

3. X-axis position is denoted by *xp* and the range of x-positions is *xpr*.

4. Y-axis position is denoted by *yp* and the range of y-positions is *ypr*.

5. X-axis stimulus size is denoted by *xs*.

6. Y-axis stimulus size is denoted by *ys.*

7. Stimulus contrast *c*; *c=1* white, *c=-1* black stimulus

8. Stimulus shape is denoted by *s*, values of *s* are following: for grating *s=1*, for vertical bar *s=2*, for horizontal bar *s=3*, for disc *s= 4*, for annulus *s=5*, for two stimuli *s=22* denotes two vertical bars, etc.

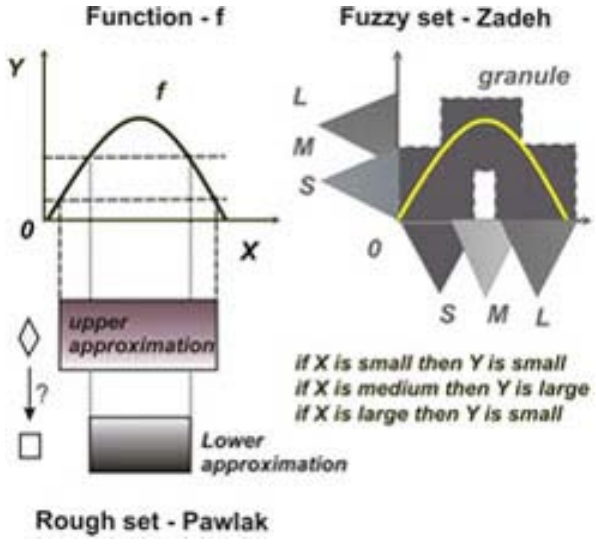

Fig. 1 Two models of the function approximation: the rough set [34] on the left, and on the right fuzzy set [68, 69] approach

Decision attributes are divided into several classes determined by the strength of the neural responses. Small cell responses are classified as class 0, medium to strong responses are classified as classes 1 to n-1 (min (n)=2), and the strongest cell responses are classified as class n. Therefore each cell divides stimuli into its own family of equivalent objects. It is similar approach to popular used in neuroscience normalization of neuronal responses from 0 to 1, but with additional values between 0 and 1.

Cell responses *(r)* are divided into n+1 range: *class 0*: activity below the threshold (e.g. 10 sp/s) labeled by r_0; class 1: activity above the threshold labeled by r_1; class n: maximum response of the cell (e.g. 100-200 sp/s) labeled by r_n.

In this paper we are using only three levels of responses: r_0, r_1, and r_2. Thus the full set of stimulus attributes is expressed as *B = {o, ob, sf, sfb, xp, xpr, yp, ypr, xs, ys, s}.*

In this work we are looking into single cell responses only in one area - V4 that will divide all patterns into equivalent (or at least similar to r degree) classes of V4-elementary granules. Neurons in V4 are sensitive only to the certain attributes of the stimulus, like for example space localization, and they are insensitive to other stimulus attribute like e.g. contrast changes. Different V4 cells have different receptive field properties, which mean that one object (B-elementary granule) can be classified in many ways by different cells (V4-elementary granules).

D. Receptive Field as a Computation Unit that Determines Similarities between Objects

Kuffler [27] first defined the receptive field as antagonistic circular center-surround filter in the output of the retina. Hubel and Wiesel [23] found elongated orientation-sensitive ON and OFF subfields in the cat primary visual cortex (V1).

Receptive field properties in the early stages of the visual pathway have been explained in terms of many different models generally as linear filters (Gaussian, Gabor or wavelets) parameterized by temporal and spatial frequencies, orientation, phase and position [4, 11]. Even if such local filters are well suited for the effective and sparse encoding of natural images, none of the computational vision systems that use them have managed to achieve robust recognition performance. It is appropriate; therefore, to consider different strategies for image processing that assist recognition.

In agreement with tuning cell's properties and graded firing rates, we assume that generally stronger neuronal responses measured in spikes/sec better classify stimulus attributes related to RF properties than weaker neuronal responses. In other words, higher response means that a certain attribute of the object and RF are more similar (with higher r) than for smaller responses.

However, we will make following modification to the classical view: we divide neuronal activity into several ranges: below a certain threshold we assume that very weak activity is not related to the stimulus (a classical approach); for activity above the threshold, as an example, we will discuss medium and strong responses in different ranges of spike frequencies (see Fig. 1).

As it is explained in Fig. 1, lower approximation (strong) neural response is related to certainty (belief □) in the classification of object attributes, whereas upper approximation (weaker) response is related to the possibility (plausibility◊) that an object may have detected attributes. Therefore our hypothesis is that by studying the strength of single cell responses to different stimulus attributes, we can find ranges of "similarities" between stimulus and RF properties. In this paper we are looking for the basis of how the brain changes the precision of object classification from uncertain to confident. Let us take a simple example like the RF of ON-center retinal ganglion cell (GC) approximated by the DOG function (like in Fig. 2). If size of our object – white spot is near the RF center diameter of GC than cell responses are larger than for smaller spots. We say that the RF better fits (more similar) to the larger spot (size attribute of the object) when GC gives stronger responses (Fig. 1: lower vs. upper approximation). Another possibility is a fuzzy set approximation (Fig. 2 right side). In this model we have three granules: small spot size give small responses, larger spot size (near size of the RF center) gives large responses, even larger spot size (that also partly covers RF surround) gives smaller responses. These two models are exchangeable but they are related to the first order, linear responses measured by mean spike frequency or by a first harmonic if the stimulus changes its intensity in time. In this case the second stimulus attribute is the optimal frequency (spatial vs. temporal frequency tuning). However, even if this example is limited to the retinal output that is not influenced by feedback from higher areas, retinal classification processes are probably more complex. By a more careful analysis of the spike train and its frequencies in response to change of the light spot diameter and frequency shows a wide range of different oscillatory responses [40, 46]. We have revealed (also in the intracellular recordings) that synchronization of certain oscillations with the stimulus might code certain stimulus attributes [46, 47]. More generally the retina (and the brain) may be seen as a system of coupled nonlinear oscillators, which synchronizations might be related to cognition [48, 67].

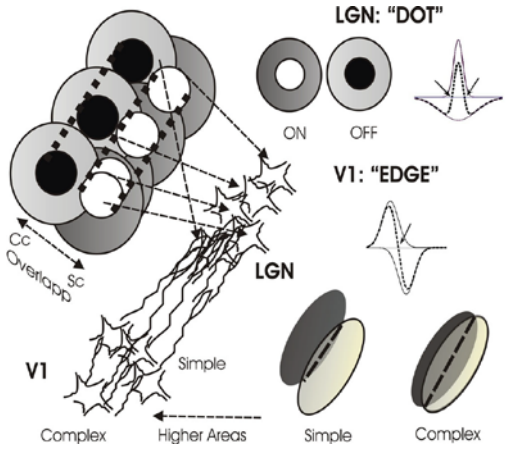

Fig. 2 Modified schematic shows RF of LGN, simple and complex V1 cells. ON- OFF-center LGN RF is well described by DOG (difference of Gaussian) functions. Aligned LGN RF may give orientation properties of V1 simple cells. V1 complex cells may arise from overlapping V1 simple cells or by higher area modulations. On-and off-subfields of V1 cells can be approximated by shifted Gaussian functions (see text)

E. Decision Rules for a Single Neuron

Each neuron in the central nervous system sums up its synaptic inputs as postsynaptic excitatory (EPSP) and inhibitory (IPSP) potentials that may cause its membrane potential to exceed the threshold and to generate an action potential. A single neuron approximate collective (thousands of interacting synapses with different weights) input information to the distributive one (unique decision in a single output). In principle, a single spike (action potential) can be seen as a result of neuronal computation (decision of the neuron), but in this work we will not take into account internal dynamics of the system and therefore we will mainly estimate neuronal activity as spikes mean frequency (as described above). This complex synaptic potential summation process is related in sensory (here only visual) systems with the receptive field properties of each neuron. Below we will show how neurons in different parts of the brain change visual information in their receptive fields into decisions (perform computations).

An extension of this approach will be to take into account membrane properties as assembly of ion channels with different dynamic. In this case the membrane can sense different frequencies in assemble of input (synaptic) signals and generate spikes with complex frequency patterns. It is the basis of the oscillatory theory of the cognition. In the retina, ganglion cells show intracellular oscillations that for certain parameters of the stimulus that can lock (see above) to the input giving appropriate burst of spikes [48]. Then the decision become more complex as the mean spike frequency give information about stimulus attributes that fit to RF properties, their frequency can give additional information about other stimulus attributes. Therefore, oscillations can be seen as a higher order decisions related to object's attributes.

F. Decision Rules for Thalamus - LGN

Each LGN cell is sensitive to luminance changes in a small part of the visual field called the receptive field (RF). The cells in LGN have the concentric center-surround shapes of their RFs, which are similar to that in the retinal ganglion cells [27]. We consider only on- and off type RFs. The on-(off) type cells increase (decrease) their activity by an increase of the light luminance in their receptive field center and/or decrease of the light luminance in the RF surround (Fig. 2).

Below are examples of the decision rules for on-off-center LGN cells with the RF position: xp_0, yp_0. We assume that there is no positive feedback from higher areas therefore the maximum response is r_1.

DR_LGN_1: $xp_0 \wedge yp_0 \wedge xs_{0.1} \wedge ys_{0.1} \wedge c_1 \wedge s_4 \text{ -> } r_1$ (1)

DR_LGN_2: $xp_0 \wedge yp_0 \wedge xs_{0.3} \wedge ys_{0.3} \wedge c_{-1} \wedge s_5 \text{ -> } r_1$ (2)

We interpret that the changes in the luminance of the light spot s4 that covers the RF center (the first rule) or annulus s5 that covers the RF surround (the second rule) gives neuronal response r1. We assume that other stimulus parameters like contrast, speed, and frequency of luminance changes, etc. are constant and optimal, and that the cell is liner and therefore we measure response of the cell activity synchronized with the stimulus changes (the first harmonic). Depending on the cell type the phase shift between stimulus and the response is near 0 or 180 deg if we do not take into account the phase shift related to the response delay. Instead using light spots (c1) or annuli (c-1) one can use a single, modulated with the drifting grating circular patch covering the classical RF. By changing the spatial frequency of the drifting grating one can stimulate only the RF center for high spatial frequencies or center and surround for lower spatial frequencies, which gives the following decision rule:

DR_LGN_3: $xp_0 \wedge yp_0 \wedge xs_{0.3} \wedge ys_{0.1} \wedge sf_{0.4} \text{ -> } r_1$ (3)

where for example: sf = 0.4 c/d stimulates RF center and surround, sf >= 1 c/d stimulates RF center only. Notice that in agreement with above rules eqs. (1-3) a single LGN cell does not differentiate between light spot, light annulus, and patch modulated with grating. All these different objects represent the same LGN-elementary granule.

G. Decision Rules for Area V1

In the primary visual cortex neurons obtain a new property: sensitivity to the stimulus orientation, which is not observed in lower areas: retina or LGN [19]. There are two cell types in area V1: simple and complex one. They can be characterized by spatial relationships between their incremental (ON) and decremental (OFF) subfields. A simple cell has in principle separated its subfields, whereas a complex cell is characterized by the overlap of its subfields. In consequence simple cells are linear (the first harmonic dominates in their responses: $F1/\ F0 > 1$), whereas complex cells are nonlinear ($F1/\ F0 < 1$). The classical V1 RF properties can be found using small flashing light spots, moving white or dark bars or gratings. We will give an example of the decision rules for the RF mapped with the moving white and dark bars [26, 44].

A moving white bar gives the following decision rule:

DR_V1_1: $o_{90} \wedge xp_i \wedge yp_0 \wedge xs_1 \wedge ys_1 \wedge c_1 \wedge s_2 \text{ -> } r_1$ (4)

The decision rule for a moving dark bar is given as:

DR_V1_2: $o_{90} \wedge xp_j \wedge yp_0 \wedge xs_1 \wedge ys_1 \wedge c_{-1} \wedge s_2 \text{ -> } r_1$ (5)

where xp_i x-position of the incremental subfield, where xp_j x-position of the decremental subfield, c_1 - stimulus contrast (1 white, -1 black), yp_0 y-position of the both subfields, xs_k, xs_1, ys_1 horizontal and vertical sizes of the RF subfields, and s_2 is a vertical bar which means that this cell is tuned to the vertical orientation (for illustration purpose we added orientation o_{90} which not necessary because the bar s_2 is vertical). We have skipped other stimulus attributes like movement velocity, direction, amplitude, etc.

For simplicity we assume that the cell is not direction sensitive, it gives the same responses to both direction of bar movement and to the dark and light bars and that cell responses are symmetric around the x middle position (xp).

An overlap index [58] is

$$OI = (0.5(xs_k+xs_l) - |xs_i-xs_j|)/(0.5(xs_k+xs_l) + |xs_i-xs_j|) \quad (6)$$

which compares sizes of increment (xs_k) and decrement (xs_l) subfields to their separation (xs_i–xs_j). After [22] if $OI < 0.3$ (non-overlapping subfields) it is the simple cell with dominating first harmonic response (linear) and r1 is the amplitude of the first harmonic. If $OI > 0.5$ (overlapping subfields), it is the complex cell with dominating F0 response (nonlinear) and r_k are changes in the mean cell activity. Hubel and Wiesel [23] have proposed that the complex cell RF is created by convergence of several simple cells in a similar ways like V1 RF properties are related to RF of LGN cells (Fig. 2). However, there are some recent experimental evidences that the nonlinearity of the complex cell RF may be related to the feedback or horizontal connections [3].

H. Decision Rules for area V4

The properties of the RFs in area V4 are more complex than that in area V1 or in LGN and in most cases they are nonlinear. It is not clear what exactly optimal stimuli for cells in V4 are, but popular hypothesis is that they V4 cells code the simple, robust shapes. Below there is an example from [36] of the decision rules for a narrow (0.4 deg) and long (4 deg) horizontal or vertical bars placed in different positions of area V4 RF:

$$\textbf{DR\_V4\_1}: o_0 \wedge ypr_m \wedge (yp_{-2.2} \vee yp_{0.15}) \wedge xs_4 \wedge ys_{0.4} \rightarrow r_2 \quad (7)$$

$$\textbf{DR\_V4\_2}: o_{90} \wedge xpr_m \wedge (xp_{-0.6} \vee xp_{1.3}) \wedge xs_{0.4} \wedge ys_4 \rightarrow r_1 \quad (8)$$

where the first rule is related to the horizontal bar o_0 and the second rule to the vertical bar (o_{90}). The horizontal bar placed narrowly in two different y-positions $yp_{-2.2}$, $yp_{0.15}$ gives strong responses (DR_V4_1), and the vertical bar placed with wide range in two different x-positions $xp_{-0.6}$, $xp_{1.3}$ gives medium responses.

I. Complex Cell Properties Determine Local Computations

As mentioned above, the default strategy for many recognition systems based on the image encoding approach is to use local filters for the transformation of image information in terms of local (Gaussian-like) gradients. These image compressions and reconstruction strategies have had such limited success in the task of the natural object recognition that it is difficult to compare them to the recognition capabilities of primates. We suggest that it may be related to different principles: primate's image recognition strategy is different from direct image encoding by band of linear filters. Therefore, we will analyze the receptive field (RF) properties of thalamic (LGN) and cortical cells in order to compare them to linear filters used in artificial systems. At first, we will show how RF properties of simple and complex cells in V1 may emerge from the LGN RFs.

The schematic in Fig. 2 demonstrates convergence of the LGN cells into V1 cells. An array of spatially aligned RFs of LGN cells may give orientation sensitivity to a V1 simple cell (SC) [23] (Fig. 2 left side). However, the origin of the area V1 complex cell (CC) RF is less clear and several hypotheses are still under debate today: 1) there is synaptic convergence of several (phase shifted) SCs on one CC [23, 38]; 2) CC properties are an effect of LGN RFs overlap [1] (Fig. 2); 3) feedback from the higher areas can change RF properties of V1 cells from simple to complex [3].

The most popular model approximates the LGN RF by the Difference of the Gaussian (DOG) function, which linearly transforms local properties of visual images (Fig. 2 right side). As mentioned above, a popular model of V1 SC and CC RFs are Gabor or Gaussian functions, which transform image linearly, whereas the electrophysiology shows that CC RFs in V1 and higher areas are nonlinear. Intracellular recordings demonstrate that there are several distinct nonlinear processes between membrane modulation and the spike generation mechanism; therefore linearity of SC RF is an exception, which depends on stimulus parameters [26]. The simple/complex cell dichotomy is also characterized by overlap between ON and OFF RF sub-regions. More precisely, ON/OFF activating regions (ARs) can be mapped with light increment/decrement (INC/DEC) bars and described as INC/DEC ARs. Recently, it has been shown that in awake monkeys, SCs are characterized by minimal overlapping (less than 30%) of the ARs, but larger group of CCs have strongly overlapping (over 50%) ARs[26]. The response of each elongated AR can be approximate by the Gaussian function [22]. If overlap is less than 30% then we can still estimate if an INC or DEC AR was stimulated and recover the input image. However, for CC with ARs overlapping more than 50%, it is not even possible to say what the stimulus polarity in the overlapping region was. Even if Shams and von der Malsburg [61] suggested that CC population responses contain sufficient information to recover the essence of images, we will concentrate on individual cells as feedback loops act on them non-uniformly [43]. Our complex cells are from the second cortical stage (layer 2+3) and not in input layer 4, which mainly integrate lower area (thalamic) input [5, 6, 14]. Therefore, mentioned above properties of CCs eliminate them as encoders, and they only can be detectors. As shown schematically in Fig. 2, larger overlap in CC RFs make CCs better edge detectors than SCs. In addition their nonlinearities help in sharpening edge detections. Moreover, the higher areas may influence the overlap of INC/DEC ARs in V1 RFs [3], as well as other RF attributes like e.g. orientation [60]. Therefore, the region of the edge detections may become variable within the RF; we call this effect the tuning of the lower areas properties to the higher areas predictions. In addition, positive feedback from higher to lower areas may regulate edge detection sensitivity [43].

In summary, CCs even from early visual areas (V1) do not encode local image features but detect attributes to which they are tuned. In consequence higher areas can only access encoded information about images in lower areas with the help of feedback pathways.

We will divide information transformation in the brain into bottom-up (BUCs) and top-down computations (TDCs). The BUCs are determined by anatomical and physiological properties of ascending pathways, whereas TDCs are related to descending pathways.

J. Local vs. Global Computations: Simplified Connections from Thalamus to Area V4; Core vs. Matrix Projections

We will demonstrate an anatomical basis of network computation that generally suggest that there are local in each

area as well as global - between areas computations with different properties (see below). This schematic is giving evidence that a popular view of a serial computations going from lower to upper anatomical areas has to be modified. There are no pure feed forward computations as all areas are strongly interconnected.

We suggest that the retina is responsible for creating preliminary hypotheses about certain features of perceived objects [45]. In one part of the Thalamus: in the Lateral Geniculate Nucleus (LGN), each hypothesis is compared with the prediction from the higher visual areas [41, 45]. If prediction and hypothesis are in agreement the decision signal is sent to the motor system to perform action [25]. This process of predictions and hypotheses is repeated in different levels of higher visual areas. In this project, we will limit our model to three hierarchical levels: LGN, V1, and V4.

In V1 we will consider three layers: L2/3, L4, and L5/6 (Fig. 3; the V1 anatomical model is based mainly on [5, 6, 14-16, 24, 30, 25, 53-55]). In the first step specific (core) input from the thalamus (LGN) activates layer 4 Pyramidal cells (L4P) and small inhibitory basket (L4B) neurons. We simplify here; in fact, parvocellular cells project to layer 4Cbeta, and the magnocellular system projects to layer 4Calpha with or without collaterals to layer 6, see schematic in [55]. Step 2: L4P cells activate local pyramidal cells (L2/3P) and basket cells (L3B) in layer 3. (For more complex details see [55]). Step 3: L2/3P pyramidal cells activate layer 5 pyramidal cells (L5P). Step 4: Layer 6 pyramidal cells (L6P) are activated, and L5P cells activate global structures of the superficial pyramidal cells that integrate information from many pyramidal cells with the help of vertically disposed double banquet cells (L2/3DB), and from the higher visual areas (dashed line). Therefore we say that in the superficial pyramidal cells the hypothesis related to the locally processed signal is compared with the predictions. In the next two steps results of the hypothesis testing are sent to L5P (Step 5), L6P (Step 6) and to thalamus (Step 7) as the second correction of the primary prediction. In Step 5, layer 6 pyramidal cells give not only feedback to the thalamus but also to cells in layer 4, and therefore correct input from the thalamus. Step 7 again starts the computation process in V1 but with multiple corrections, as described above.

After preliminary computation in V1, activity from L3P is fed forward to the higher areas; in our model we simplify such that V1 makes direct projection to the area V4 (Fig. 3). Notice those computations in higher areas, here only in V4, run almost in parallel to V1. After two steps: LGN to layer 4 and L4P to superficial layers (see two-step local flow of information in [15]) in V1 local activity is sent to higher areas (Fig. 3). In the next two steps superficial cells in V1 integrate activity from other L2/3P, and L5P V1 cells, and from higher areas (L6P V4 cells) and send corrected information to V4. But during this time preliminary computation in V4 was performed in two-steps: from V1 to V4 layer 4 cells and to L2/3P cells.

In V4 the computation process runs in a similar order to that in V1 but neuronal density is lower than in V1 and receptive fields are about 4 times larger so we assume that each input neuron in V4 integrates information from four output V1 cells. The main functional difference is that V4 superficial layer cells activate V1 L2/3 cells, and V4 L6P cells activate V1 L2/3 cells and also directly activate LGN cells.

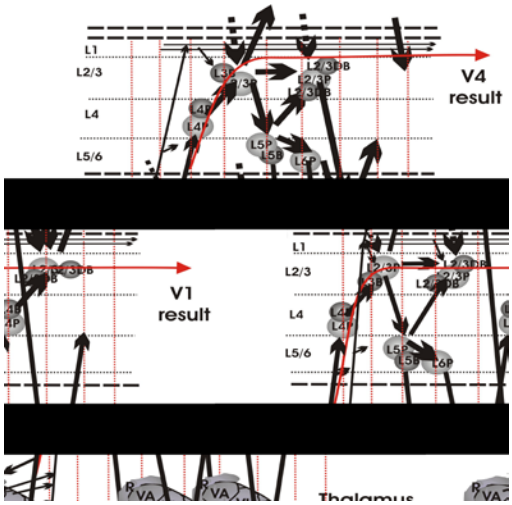

Fig. 3 Simplified schematic ascending connections from the thalamus to area V1 and V4. Ascending computations interact with feedbacks from higher areas that are faster (myelinated fibers) that slow local computations.

Our model takes into account feed forward and feedback interactions between local and global computations and activities. Local interactions are intra-layer and between layers, and global interactions are between different areas. Therefore we will give computational meaning to different substructures (layers) and we will get physiological meaning of our partial (layer or area related) computational results. However, our main purpose is to show that the computational performance of the whole system can follow our [37] and others' [21, 31, 32] experimental and theoretical results describing the complexity of the V4 receptive field, and can outperform others' object recognition system models.

It is now well established that every nucleus in the dorsal thalamus receives input from the cortex and projects into it. The classical organization (core-projection) arises in the LGN (lateral geniculate nucleus) for the visual system (from medial geniculate, and ventral posterior nuclei for other sensory systems) and focus on layer IV and less on layers III and VI. Many other thalamic nuclei projects (matrix-projection) upon several cortical areas terminate primarily in the superficial layers (I and II). Thalamic relay cells get 44% synapses from cortex, 16% from the retina and 40% inhibitory synapses from the reticular nucleus (RN) and interneurons (5%) [15, 25]. 70% synapses in RN are corticothalamic collaterals, some of them are from thalamocortical collaterals. Termination of the axons of matrix cells in superficial layers on the apical dendritic sprays of these cells may set up a coincident detector to the core projection. The spread of activity across assemblies may occur by feedback projections from layer V to a new thalamic nuclei and horizontal connections of matrix cells. Corticothalamic cells with somas in layer V have far more extensive axonal ramifications in the cortex and thalamus. They have dendrites in the layer I and their axons give off a number of horizontal collaterals in layers III and V and then descend to the thalamus and to other subcortical structures such as the tectum, other parts of the brain stem, or the spinal cord. Unlike the axons of a layer VI cells, axons of layer V cells do not give off collaterals to the reticular nucleus and they are not restricted to the nucleus from which their parent cortical area receives inputs (like for a layer VI neurons). Their axons extend into one or more adjacent nuclei, although in each nucleus the terminals can be more focused than those of the axons of layer VI cells. The focusing of the layer V projection in comparison with layer VI projection

does not imply a greater degree of topographic specificity because their intracortical projections are widespread in comparison to highly columnar layer VI projections.

K. Logic of the Anatomical Connections

As it was mentioned above, our model consists of three interconnected visual areas. Their connections can be divided into feedforward (FF) and feedback (FB) pathways. We have proposed [45] that FF connections are related to the hypothesis about stimulus attributes and FB are related to predictions. Below, we suggest that the different anatomical properties of the FB and FF pathways may determine their different logical rules.

We define LGN_i, as LGN *i-cell* attributes for cells $i=1,\ldots,n$, $V1_j$ as primary visual cortex *j-cell* attributes for cells $j=1,\ldots,m$, and $V4_k$ as area V4 attributes for cells $k=1,\ldots,l$.

The specific stimulus attributes for a single cell can be found in the neurophysiological experiment by recording cell responses to the set of various test stimuli. As we have mentioned above, cell responses are divided into several (here 3) ranges, which will define several granules for each cell. It is different from the classical receptive field definition, which assumes that the cell responds (logical value 1) or does not respond (logical value 0) to the stimulus with certain attributes. In the classical electrophysiological approach all receptive field granules are crisp. In our approach, cell responses below the threshold – r_0, have logical value *0*, the maximum cell responses - r_2, have a logical value *1* but we will introduce cell responses between r_0 and r_2, in this paper only one value r_1. The physiological interpretation of cell responses between the threshold and the maximum response may be related to the influence of the feedback, horizontal pathways or matrix projections. We assume that the tuning of each structure is different and we will look for decision rules in each level that give responses r_1 and r_2. For example, we assume that r_1 means that the local structure is tuned to the attributes of the stimulus and such granule for *j-cell* in area V1 will be define as $[u]_{1V1j}$.

1) Bottom-Up Computations (BUCs):

We will describe the logic of BUCs on the basis of LGN to V1 pathways, and by simplified direct and indirect influence of area V1 on area V4. Thalamic axons target specific cells in layers 4 and 6 of the primary visual cortex (V1). As Hubel and Wiesel [23] proposed, LGN cells determine orientation of SCs with their receptive fields arranged along the preferred orientation of the V1 cell (Fig. 2). There is high specificity between RF properties of the LGN cells and SC if they have monosynaptic connections [1]. The precision goes beyond simple retinotopy and includes such RF properties as RF sign, timing, subregion's strength, and size [1]. This high specificity of connections determines that V1 cell response is a result of assembly activity of several specific LGN cells "connected" by the logical "AND" (" $\wedge$ ") as it was already discussed above. This is related to the fact that several aligned receptive fields in LGN must be simultaneously activated ("and") in order to activate V1 cell connected to them [23]. As Sherman and Guillery [62] proposed, we will call such inputs drivers. We can write formally this proposal as follows:

DR_LGN_V1:

$$r^{LGN}(x_0, y_0) \wedge r^{LGN}(x_1, y_1) \wedge \ldots \wedge r^{LGN}(x_n, y_n) \text{->} r^{V1}(x_k, y_k) \quad (9)$$

We understand it as the decision rule (DR) how cells from LGN influence activity of the V1 neuron (DR_LGN_V1). This rule describes response $r^{V1}(x_k, y_k)$ of the area V1 cell with coordinates (x_k, y_k) as determined by responses $r^{LGN}(x_i, y_i)$ of n+1 LGN cells with coordinates (x_0, y_0) to (x_n, y_n) monosynaptic connections to a V1 cell. From this rule we may infer that all monosynaptic inputs from the LGN must have sufficient strength in order to obtain significant V1 response. At this stage we propose that SCs and CCs have similar decision rules, but if LGN cells are not directly connected to V1 CC then synaptic weight will be "effective" synaptic weight.

Similar rules apply for the BUCs in the higher areas. There are relatively small direct connections from V1 to V4, but we also take into account V1 to V2 [55] and V2 to V4 feed forward connections. These connections are highly organized but variable, especially in V4 [53]. We assume that, direct or indirect connections from area V1 to V4 provide driver inputs, which fulfill similar principles as connections from the LGN to V1, and implement the following decision rules:

DR_V1_V4:

$$r^{V1}(x_0, y_0) \wedge r^{V1}(x_1, y_1) \wedge \ldots \wedge r^{V1}(x_n, y_n) \text{->} r^{V4}(x_k, y_k) \quad (10)$$

We assume that the neuron in area V4 receives driver inputs directly from cells in area V1 as well as indirectly through area V2 with highly specific RF properties (as described above for connections between LGN and V1 – equation 1). Therefore, the logical "and" has the same meaning as above: every input neuron from V1 "connected" to V4 (x_n, y_n) cell must be activated in order to activate V4 cell (more explicit formula is the appendix). However, in this case "connection" can be changed by the descending pathways (see below).

2) Top-Down Computations (TDCs):

The bases of TDCs are anatomical and physiological properties of descending pathways. Their function is to perform similarity verification that may lead to recognition. In the primate visual system the first descending pathway is from area V1 to the LGN.

Experimental results show that V1 feedback connections are restricted to the LGN region, visual-topically coextensive with the size of the classical RF of V1 layer 6 cells [2]. We will call feedback inputs modulators [62] with the following decision rule:

DR_V1_LGN:

$$r^{LGN}(x_i, y_i) * r^{V1}(x_1, y_1) \vee r^{LGN}(x_i, y_i) * r^{V1}(x_2, y_2) \vee \ldots \vee r^{LGN}(x_i, y_i) * r^{V1}(x_k, y_k) \text{->} r^{LGN}(x_i, y_i) \quad (11)$$

this rule says that when the activity of a particular V1 cell $r^{V1}(x_j, y_j)$ is in agreement with activity of the LGN cells $r^{LGN}(x_i, y_i)$ it response will multiplicatively ("*")increase, where (x_j, y_j) are coordinates of V1 cells from index $i=1,\ldots,k$, which have anatomical connections with the LGN cell with coordinates (x_i, y_i) [43]. The logical "or" ("v") is related to tuning of different LGN cells in agreement with preferred RF property of the V1 cell.

Decision Rules for TDCs from V4 to V1 or V4 to LGN will have similar syntax even if anatomical and physiological properties of the feedback pathways are different. Retrograde anatomical tracing has shown descending axons from area V4

directly to area V1 [54]. Axons of V4 cells span into area V1 in distinct clusters or in a linear array. The different semantics in decision rules are V4 cell specific and are related to the shapes of individual and variable axon branches in area V1. An axon's cluster that has terminals on V1 cells near "pinwheel centers", (where cells show sub-threshold responses to all orientations [59] – will be responsible for the V4 subfield orientation tuning. If a linear array of terminals is connected to V1 neurons with similar orientation preference (narrowly tuned neurons [59]) – place tuning will take place. Retrograde tracing from area V4 showed axons projecting to different layers of the LGN with terminations in distinct clusters or in linear branches. These projections will also tune orientation and place of V4 cell subfields but with different precision than V4 to V1 pathways.

To summarize, object recognition has two stages: at first BUCs classify all possible objects' similarities in different visual areas; in the next stage TDCs verify BUCs classification. In the following paragraph we will apply our computational model to experimental data from the area V4.

III. RESULTS

We have analyzed the experimental data from several neurons recorded in the monkey's V4 [37]. Below we show a modified figure from the above work (Fig.1), along with the associated decision table (Table Ⅰ). On the basis of the decision table we have made a schematic of the optimal stimulus for this cell (Fig. 4 right side). Fig. 4 (left side) shows the cell's responses to the stimulus, which was a long narrow bar with vertical (Fig. 4 C) or horizontal (Fig. 4 D) orientation.

The decision table (Table 1) describes properties of stimuli and their position as a function of response strength. This table is converted into a schematic (right of Fig. 1), which shows areas of cell responses related to category 1 (upper part) and to category 2 (lower part). Strong cell responses are not symmetric along the middle of the receptive field, but divide the receptive field into several smaller subfields.

These results are the basis of the idea that the receptive field of V4 neurons can be divided into several independent parts. Our results can be presented as follows:

Decision rules:

$$\textbf{\textit{DR1}}: o_{90} \wedge (xpr_{0.5} \vee xpr_{0.6}) \wedge xs_{0.4} \wedge ys_4 \text{->} r_2 \quad (12)$$

$$\textbf{\textit{DR2}}: o_0 \wedge (ypr_{1.2} \vee ypr_{0.7}) \wedge xs_4 \wedge ys_{0.4} \text{->} r_2 \quad (13)$$

TABLE I DECISION TABLE FOR THE CELL SHOWN IN FIG.4 ATTRIBUTES ob, sf, sfb WERE CONSTANT AND ARE NOT PRESENTED IN THE TABLENOT PRESENTED IN THE TABLE

Cell	o	xp	xpr	yp	ypr	xs	ys	s	r
12a	90	-0.6	1.2	0	0	0.4	4	2	1
12a1	90	-0.6	0.6	0	0	0.4	4	2	2
12a2	90	1.3	1	0	0	0.4	4	2	1
12a3	90	1.3	0.5	0	0	0.4	4	2	2
12b	0	0	0	-2.2	1.6	4	0.4	3	1
12b1	0	0	0	-2.2	1.2	4	0.4	3	2
12b2	0	0	0	0.15	1.3	4	0.4	3	1
12b3	0	0	0	0.15	0.7	4	0.4	3	2

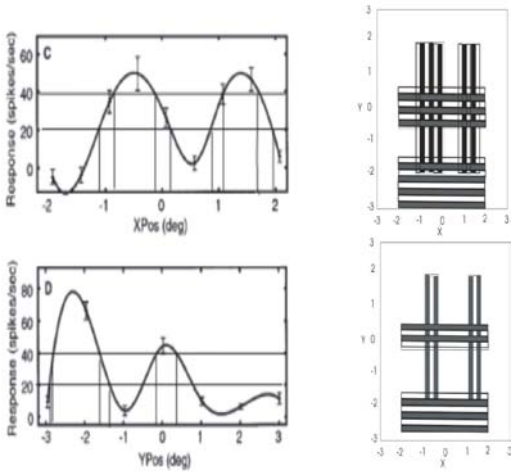

Fig. 4 Curves represent approximated responses of a cell from area V4 to vertical (C), and horizontal (D) bars. Bars change their position along x-axis (Xpos) or along y-axis (Ypos). Responses of the cell are measured in spikes/sec. Mean cell responses ± SE are marked in the figures. Cell responses are divided into three ranges (concepts) by two horizontal lines. On the right is a schematic representation of cell response on the basis of Table I. Vertical and horizontal bars in certain x- and y-positions gave strong (r1: class 1 – upper schematic) or very strong (r2: class 2 – lower schematic) responses.

Notice that Figs. 4 and 5 show possible configurations of the optimal stimulus. However, they do not take into account interactions between several stimuli, when more than one subfield is stimulated.

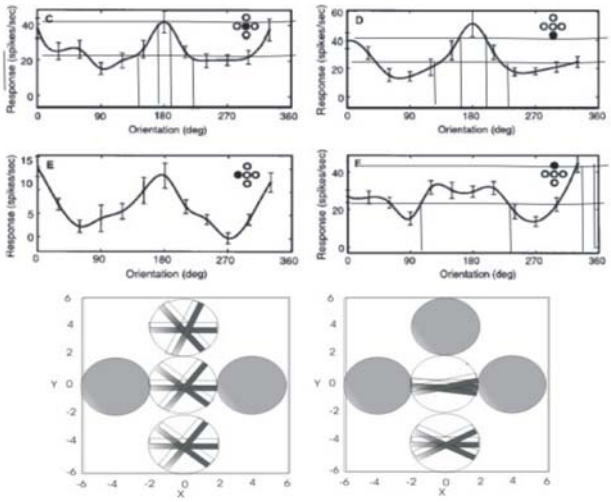

Fig. 5 Modified plots on the basis of [37] (upper plots), and their representation on the basis of table 2 (lower plots). C-F Curves represent responses to different orientations of one V4 cell when its subfields (their positions are shown in plots) are covered with a 2 degree grating discs 2 degrees apart in a 6 degree receptive field. Lower plots: Gray circles indicate cell response below 20 spikes/s. Plots on the left are related to r1: class 1, and plots on the right to r2: class 2 responses.

In addition there are Subfield Interaction Rules:

SIR1: facilitation when stimulus consists of multiple bars with small distances (0.5-1deg) between them, and inhibition when distance between bars is 1.5 -2 deg.

SIR2: inhibition when stimulus consists of multiple similar discs with distance between them ranging from 0 deg (touching) to 3 deg [36].

SIR3: centre-surround interaction, which was described earlier [42] in detail.

Let us simplify the category $0 < ob < 50$ and denote it as ob_n (narrow orientation bandwidth), $ob > 100$ as ob_w (wide orientation bandwidth), $0 < sfb < 2$ as sfb_n, and $sfb > 2.5$ as sfb_w.
The Decision rules are as follows:

$$\textbf{DR3}: ob_n \wedge (yp_0 \vee yp_2) \rightarrow r_2, \quad (14)$$
$$\textbf{DR4}: ob_w \wedge xp_0 \rightarrow r_1, \quad (15)$$
$$\textbf{DR5}: sfb_n \wedge yp_0 \rightarrow r_2, \quad (16)$$
$$\textbf{DR6}: sfb_w \wedge xp_0 \rightarrow r_1, \quad (17)$$

TABLE II DECISION TABLE FOR ONE CELL SHOWN IN FIG.5 ATTRIBUTES xpr, ypr, s ARE CONSTANT AND ARE NOT PRESENTED IN THE TABLE

Cell	o	ob	sf	sfb	xp	yp	r
3c	172	105	2	0	0	0	1
3c1	10	140	2	0	0	0	1
3c2	180	20	2	0	0	0	2
3d	172	105	2	0	0	-2	1
3d1	5	100	2	0	0	-2	1
3d2	180	50	2	0	0	-2	2
3e	180	0	2	0	-2	0	0
3f	170	100	2	0	0	2	1
3f1	10	140	2	0	0	2	1
3f2	333	16	2	0	0	2	2
5a	180	0	2.3	2.6	0	-2	1
5b	180	0	2.5	3	0	2	1
5c	180	0	2.45	2.9	0	0	1
5c1	180	0	2.3	1.8	0	0	2

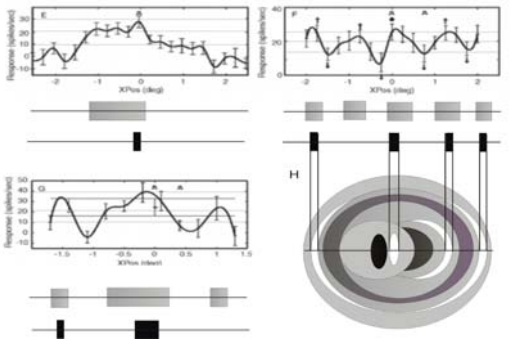

Fig. 6 Modified plots from [37]. Curves represent responses of two cells from area V4 to small single (E) and double (F, G) vertical bars. Bars change their position along xaxis (Xpos). Responses are measured in spikes/sec. Mean cell responses SE are marked in E, F, and G. Cell responses are divided into three ranges by thin horizontal lines. Below each plot are schematics showing bar positions giving r1 (gray) and r2 (black) responses; below (E) for a single bar, below (F and G) for double bars (one bar was always in position 0). (H) This schematic extends responses for horizontally placed bars (E) to the whole RF: white colour shows excitatory related to r2 responses, gray color is related to r1 responses and black color inhibitory interactions between bars

Below we give an example of the SIR1. We will analyze experiments where the RF is stimulated at first with a single small vertical bar and later with two bars changing their horizontal positions. One example of V4 cell responses to thin (0.25 deg) vertical bars in different horizontal positions is shown in the upper left part of Fig. 6 (Fig. 6E). Cell response has maximum amplitude for the middle ($XPos = 0$) bar position along the x-axis. Cell responses are not symmetrical around 0. In Fig. 6F, the same cell (cell 61 in table 3) is tested with two bars. The first bar stays at the 0 position, while the second bar changes its position along x-axis. Cell responses show several maxima dividing the receptive field into four areas. However, this is not always the case as responses to two bars in another cell (cell 62 in table 3) show only two minima (Fig. 6G). Horizontal lines in plots of both figures divide cell responses into the three categories r_0, r_1, r_2, which are related to the mean response frequency (see Methods). Stimuli attributes and cell responses classified into categories are shown in Table 3 for cells in Fig. 6.

TABLE III DECISION TABLE FOR ONE CELL SHOWN IN FIG.6 ATTRIBUTES o, ob, sf, sfb ARE CONSTANT AND ARE NOT PRESENTED IN THE TABLE

Cell	xp	xpr	xs	ys	s	r
61e	-0.7	1.4	.25	4	2	1
61f1	-1.9	0.2	.25	4	22	2
61f2	0.1	0.2	.25	4	12	2
61f3	1.5	0.1	.25	4	22	2
61f4	-1.8	0.6	.25	4	22	1
61f5	-0.4	0.8	.25	4	22	1
61f6	0.4	0.8	.25	4	22	1
61f7	1.2	0.8	.25	4	22	1
62g1	-1.5	0.1	.25	4	22	2
62g2	-.15	0.5	.25	4	22	2
62g3	-1.5	0.6	.25	4	22	1
62g4	-.25	1.3	.25	4	22	1
62g5	1	0.6	.25	4	22	1
62h1	-0.5	0	.5	1	44	2
62h2	1	1	1	1	44	1
62h3	0.2	0.1	.25	4	22	2

We assign the narrow (xpr_n), medium (xpr_m), and wide (xpr_w) x position ranges as follows: xpr_n if ($xpr: 0<xpr <= 0.6$), medium xpr_m if ($xpr: 0.6 <xpr <=1.2$), wide xpr_w if ($xpr: xpr>1.2$). We assign the narrow (ypr_n), medium (ypr_m) and wide (ypr_w) y position range: ypr_n if ($ypr: 0<ypr <= 1.2$), medium ypr_m if ($ypr: 1.2 <ypr <=1.6$), wide ypr_w if ($ypr: ypr>1.6$).

On the basis of Fig. 6 and the decision table 3 (also compare with [37]) the one-bar study can be presented as the following decision rules:

$$\textbf{DR\_V4\_5}: o_{90} \wedge xpr_n \wedge xp_{0.1} \wedge xs_{0.25} \wedge ys_{0.4} \rightarrow r_2 \quad (18)$$

$$\textbf{DR\_V4\_6}: o_{90} \wedge xpr_w \wedge xp_{-0.2} \wedge xs_{0.25} \wedge ys_{0.4} \rightarrow r_1 \quad (19)$$

We interpret these rules that r_1 response in eq. (18) does not effectively involve the feedback to the lower areas: V1 and LGN. The descending V4 axons have excitatory synapses not only on relay cells in LGN and pyramidal cells in V1, but also on inhibitory interneurons in LGN and inhibitory double banquet cells in layer 2/3 of V1. As an effect of the feedback, only narrow range of area V4 RF responded with a high r_2 activity to a single bar stimulus, whereas in outside area excitatory and inhibitory feedback influences compensate each other.

Decision Rules of Two-bar (DRT):

DRT1:

$$o_{90} \wedge xpr_n \wedge ((xp_{-1.9} \vee xp_{0.1} \vee xp_{1.5}) \wedge xs_{0.25} \wedge ys_4)_1 \wedge (o_{90} \wedge xp_0 \wedge xs_{0.25} \wedge ys_4)_0 \rightarrow r_2 \quad (20)$$

DRT2:

$$o_{90} \wedge xpr_m \wedge ((xp_{-1.8} \vee xp_{-0.8} \vee xp_{0.4} \vee xp_{1.2}) \wedge xs_{0.25} \wedge ys_4)_1 \wedge (o_{90} \wedge xp_0 \wedge xs_{0.25} \wedge ys_4)_0 \rightarrow r_1 \quad (21)$$

Two-bar decision rules claim that: the cell's responses to two bars are strong if one bar is in the middle of the RF (bar with index 0 in decision rules) and the second narrow bar (bar with index 1 in decision rules) is in the certain positions of the RF eq. (20). But when the second bar has medium width the max cell responses became weaker eq. (21). Responses of other cells are sensitive to other bar positions (Fig. 6G). These differences could be correlated with anatomical variability of connections especially of the descending axons. As mentioned above V4 axons in V1 have distinct clusters or linear branches. Descending pathways are modulators and therefore their rules contain logical or which consequence is that not all excitatory areas become more active as a result of the feedback.

IV. DISCUSSION

In this paper we have considered possible mechanisms on "how visual system can figure out" properties of the unseen object. We have proposed to formalize the receptive field (RF) properties with help of rough and fuzzy set theories. By using this concept and by normalization to several levels neuronal responses one can check decisions performed by each neuron in response to different stimuli. These decisions tell us how similar RF and object (stimulus) properties are.

Neurons in area V4 integrate an object's attributes from the properties of its parts in two ways: (1) within the area via horizontal or intra-laminar local excitatory-inhibitory interactions, (2) between areas via feedback connections tuned to lower visual areas. Our research put more emphasis on feedback connections because they are probably faster than horizontal interactions [21]. Different neurons have different Subfield Interactions Rules as described in the Results section and perceive objects by way of multiple "unsharp windows". If an object's attributes fit the unsharp window, a neuron sends positive feedback [43] to lower areas, which as described above, use "modulator-type" logic (MTL) to sharpen the attribute-extracting window and therefore change response of the neuron from class 1 to class 2. The above analysis of our experimental data leads us to suggest that the central nervous system chiefly uses at least two different logical rules: "driver-type" logical (DTL) rule" and "modulator logical (MTL) rule." The first, DTL processes data using a large number of possible algorithms (over-representation). The second, MTL supervises decisions and chooses the right algorithm. As we have described, there are experimental findings [3, 5] suggesting that properties of RF in lower areas can be tuned by descending pathways. These findings are basis for the universality of our visual system that by learning and trials can recognize unseen objects by changing hypotheses about their actual properties. It is based on similarities. Other proposed matching shape similarities an array of multi-scale, multi-oriented "Gabor-jets" detectors [28]. Problem with such models, that can achieve good accuracy in e.g. face recognition, is that they are not sensitive to contour variations that are very important in the object recognition.

Physical properties of objects are different from their psychological representation. Gärdenfors [20] proposed to describe the principle of human perceptual system as grouping objects by similarities in the conceptual space. Human perceptual systems group together similar objects with unsharp boundaries [20], which means that objects are related to their parts by rough inclusion or that different parts belong to objects with some approximation (degree) [39]. We suggest that similarity relations between objects and their parts are related the hierarchical relationships between different visual areas. These similarities may be related to resonance [10] or synchronizations of multi-resolution, parallel computations and are difficult to simulate using a digital computer [48].

Treisman [65] proposed that our brains extract features related to different objects using two different procedures: parallel and serial processing. The "basic features" were identified in psychophysical experiments as elementary features that can be extracted in parallel. Evidence of parallel features extraction comes from experiments showing that the extraction time becomes independent of the number of objects. Other features need serial searches, so that the extraction time is proportional to the number of objects. High-level serial processing is associated with integration and consolidation of items combined with conscious awareness. Other low-level parallel processes are rapid, global, related to high-efficiency categorization of items and largely unconscious [65]. Treisman [65] showed that instances of a disjunctive set of at least four basic features could be detected through parallel processing. Other researchers have provided evidence for parallel detection of more complex features, such as shape from shading [50] or experience-based learning of features of intermediate complexity [66].

Thorpe et al. [64] found that human and non-human primates can rapidly and accurate categorize of briefly flashed natural images. Human and monkey observers are very good at deciding whether or not a novel image contains an animal even when more than one image is presented simultaneously [56]. The underlying visual processing reflecting the decision that a target was present is less than 150 ms [54]. These findings are in contradiction to the classical view that only simple, "basic features," likely related to early visual areas like V1 and V2, are processed in parallel [65]. Certainly, natural scenes contain more complex stimuli than "simple" geometric shapes. It seems that the conventional, two-stage perception processing model needs correction, because to the "basic features" we must add a set of unknown intermediate features. We propose that at least some intermediate features are related to receptive field properties in area V4. Area V4 has been associated with shape processing because its neurons respond to shapes [13] and because lesions in this area disrupt shape discrimination, complex-grouping discriminations [31], multiple viewpoint shape discriminations [32], and rotated shape discriminations [22]. Although Thorpe et al. [64] assumed that object recognition in prefrontal cortex was done on a feed forward, one-pass basis, and our anatomical schematic (Fig. 3) shows that there is a fast, global feed forward-feedback pathway that interacts with performed in parallel local computations. As an example of such computations are recordings in IT where neurons respond at first to an object's information in a coarser scale (gender of the face) and later to finer details (from global context to detailed information) [63].

As it was mentioned in the introduction, our model has some similarities to predictive coding models [12, 29, 51]. The most popular model [29] is based on the Bayes' rule and introduce hierarchical Bayesian inference model in the visual cortex. This model assumes that each visual area is influenced mainly by its direct neighbors (concept of Markov chain) and maximizes by competition the probability of its computed features [29]. These assumptions have no physiological basis, as there are also top-down connections omitting neighbors

like for example area V4 has direct connections to V1 [54] or V4 to LGN [49]. Another problem with this model is related with different and still not very clear roles of core and matrix projections (see Methods: II.J section). Also such local rules may lead to many iterations and long computation time. In general, inference grounded on the Bayes' rule assumes that some "prior" probability (knowledge) without knowledge about data is given first. When data is obtained, posterior probability is computed. Then it is used to verify the prior probability. In the rough set model the lower and the upper approximation of a set, computed directly from the data, satisfy the Bayes' rule without referring to subjective prior and posterior probabilities [35].

By applying rough sets to V4 neuron responses, we have differentiated between bottom-up information (hypothesis testing) related to the sensory input, and predictions, some of which can be learned but are generally related to positive feedback from higher areas. If a prediction is in agreement with a hypothesis, object classification will change from category 1 to category 2. Our research suggests that such decisions can be made very effectively during pre-attentive, parallel processing in multiple visual areas. In addition, we found that the decision rules of different neurons can be inconsistent.

One should take into account that modeling complex phenomena details the use of local models (captured by local agents, if one would like to use the multi-agent terminology [57]) that should be fused afterwards. This process involves negotiations between agents [57] to resolve contradictions and conflicts in local modeling. One of the possible approaches in developing methods for complex concept approximations can be based on the layered learning [36]. Inducing concept approximation should be developed hierarchically starting from concepts that can be directly approximated using sensor measurements toward complex target concepts related to perception. This general idea can be realized using additional domain knowledge represented in natural language.

We have proposed decision rules for different visual areas and for FF and FB connections between them. However in processing our V4 experimental data, we also have found inconsistent decision rules. These inconsistencies could help process different aspects of the properties of complex objects. The principle is similar to that observed in the orientation tuning cells of the primary visual cortex. Neurons in V1 with overlapping receptive fields show different preferred orientations. It is assumed that this overlap helps extract local orientations in different parts of an object. However, it is still not clear which cell will dominate if several cells with overlapping receptive fields are tuned to different attributes of a stimulus. Most models assume the "winner takes all" strategy, meaning that using a convergence (synaptic weighted averaging) mechanism, the most dominant cells will take control over other cells, and less represented features will be lost. This approach is equivalent to the two-valued ("true-false") logic implementation. Our finding from area V4 seems to support a different strategy than the "winner takes all" approach. It seems that different features are processed in parallel and then compared with the initial hypothesis in higher visual areas. We think that descending pathways play a major role in this verification process. At first, the activity of a single cell is compared with the feedback modulator by logical conjunction to avoid hallucinations. Next, the global, logical disjunction ("modulators") operation allows the brain to choose a preferred pattern from the activities of different cells. This process of choosing the right pattern may have strong anatomical basis because individual axons have variable and complex terminal shapes, facilitate some regions and features against other so called salient features. Learning can probably modify the synaptic weights of the feedback boutons, fine-tuning the modulatory effects of feedback.

CONCLUSIONS

By applying the rough set theory to neuro-physiological data we have demonstrated a new formalized approach: how the visual brain may perform object categorization in the psychophysical space. These processes are related to anatomical and physiological properties of the visual system: ascending and descending pathways are related to hypotheses and predictions and mirrored by different logical systems (DLT vs. MLT: driver-type vs. modulatory-type logic). These different logical rules look for similarities between properties of the object or its parts in comparison to RF properties of neurons in LGN, V1, V2, V4 and higher areas in the ventral stream. In agreement with previous experiences the right hypothesis that is the most similar to our predictions about the object is chosen. It is the basis of the cognition related to the first order processes (spike rates). Using the same fuzzy logical systems (DLT vs. MLT) one can describe higher order processes related to oscillatory processes. By extending of our retina model as the coupled nonlinear oscillatory system to higher visual areas we propose that in this case also DLT vs. MLT interactions will be the basis of cognition. The bottom-up system consists of a large number of possible orbits and only some of them are chosen by top-down parametric control of the lower level oscillators.

ACKNOWLEDGMENT

This work was partly supported by NIH #1R43NS064640-01 and by National Center for Research and Development # 2892/B/T02/2011/40.

REFERENCES

[1] Alonso, J. M., Usrey, W. M., & Reid, R. C. (2001). Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. Journal of Neuroscience, 21, 4002–4015.

[2] Angelucci, A., & Sainsbury, K. (2006). Contribution of feedforward thalamic afferents and corticogeniculate feedback to the spatial summation area of macaque V1 and LGN. Journal of Comparative Neurology, 498, 330–351.

[3] Bardy, C., Huang, J. Y., Wang, C., Fitz Gibbon, T., & Dreher, B. (2006). Simplification of responses of complex cells in cat striate cortex: Suppressive surrounds and 'feedback' inactivation. Journal of Physiology, 574, 731–750.

[4] Barlow, H. B. 1961 The coding of sensory messages. in Current Problems in Animal Behaviour. W. H. Thorpe, O. L. Zangwill Eds. Cambridge Univ. Press, Cambridge, pp 331–60.

[5] Binzegger T, Douglas RJ, Martin KA. A quantitative map of the circuit of cat primary visual cortex. J Neurosci. 24:8441-53, 2004.

[6] Briggs, F. and Callaway, E. M., Laminar patterns of local excitatory input to layer 5 neurons in macaque primary visual cortex, Cerebral Cortex, vol. 15 pp. 479-488, 2005.

[7] Bullier J. Integrated model of visual processing. (2001) Brain Res Brain Res Rev. 36(2-3): 96-107

[8] Bullier J, Hupé JM, James AC, Girard P. (2001) The role of feedback connections in shaping the responses of visual cortical neurons.Prog Brain Res. 2001;134:193-204.

[9] Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T. (2007) A model of V4 shape selectivity and invariance. J Neurophysiol. 98(3): 1733-50.

[10] Carpenter GA, Grossberg S, Rosen DB (1991) Fuzzy ART: fast stable

learning and categorization of analog patterns by an adaptive resonance system. Neural Networks 4:759-771.

[11] Daugman, J. D. 1985 Uncertainty relation for the resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Sci Am A. 2, 1160-1169.

[12] Dayan P, Hinton GE, Neal RM, Zemel RS. (1995) The Helmholtz machine. Neural Comput. 7(5):889-904.

[13] David, S. V., Hayden, B. Y., & Gallant, J. L. (2006). Spectral receptive field properties explain shape selectivity in area V4. Journal of Neurophysiology, 96, 3492–3505.

[14] Douglas, R. J. and Martin, K. A., Neuronal circuits of the neocortex, Annu Rev Neurosci, 27 pp. 419-451, 2004.

[15] Douglas RJ, Martin KA. Mapping the matrix: the ways of neocortex. Neuron. 56: 226-38, 2007.

[16] Fitzpatrick, D., Usrey, W. M., Schofield, B. R., & Einstein, G. (1994). The sublaminar organization of corticogeniculate neurons in layer 6 of macaque striate cortex. Visual Neuroscience, 11, 307–315.

[17] Fukushima K Cognitron: a self-organizing multilayered neural network. Biol Cybern. 1975 20: 121-36.

[18] Fukushima K Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980;36(4):193-202.

[19] Gaska JP, Pollen DA, Cavanagh P. Diversity of complex cell responses to even- and odd-symmetric luminance profiles in the visual cortex of the cat.Exp Brain Res. 1987; 68:249-59.

[20] Gärdenfors, P. (2000). Conceptual spaces. Cambridge, MA: MIT Press.

[21] Girard, P., Hupe, L, J. M., & Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. Journal of Neurophysiology, 85, 1328–1331.

[22] Girard, P., Lomber, S. G., & Bullier, J. (2002). Shape discrimination deficits during reversible deactivation of area V4 in the macaque monkey. Cerebral Cortex, 12, 1146–1156.

[23] Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology, 160, 106–154.

[24] Ichida, J. M., & Casagrande, V. A. (2002). Organization of the feedback pathway from striate cortex (V1) to the lateral geniculate nucleus (LGN) in the owl monkey (Aotus trivirgatus). Journal of Comparative Neurology, 454, 272–283.

[25] E. G. Jones Synchrony in the Interconnected Circuitry of the Thalamus and Cerebral Cortex Annals of the New York Academy of Sciences, 1157, 10–23, 2009.

[26] Kagan, I., Gur, M., & Snodderly, D. M. (2002). Spatial organization of receptive fields of V1 neurons of alert monkeys: comparison with responses to gratings. Journal of Neurophysiology, 88, 2557–2574.

[27] Kuffler, S. W. (1952). Neurons in the retina; organization, inhibition and excitation problems. Cold Spring Harbor Symposia on Quantitative Biology, 17, 281–292.

[28] Lades, M., Vortbrueggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wuertz, R. P., et al. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers, 42, 300–311.

[29] Lee TS, Mumford D. (2003) Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am A Opt Image Sci Vis. 20(7):1434-48.

[30] Lund, J. S., Lund, R. D., Hendrickson, A. E., Bunt, A. H., & Fuchs, A. F.(1975). The origin of efferent pathways from the primary visual cortex,area 17, of the macaque monkey as shown by retrograde transport of horseradish peroxidase. Journal of Comparative Neurology, 164, 287–303.

[31] Merigan, W. H. (2000). Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. Visual Neuroscience, 17, 949–958.

[32] Merigan, W. H., & Pham, H. A. (1998). V4 lesions in macaques affect both single- and multiple-viewpoint shape discriminations. Visual Neuroscience, 15, 359–367.

[33] Olshausen BA, and Field DJ. (1997). Sparse Coding with an Over complete Basis Set: A Strategy Employed by V1? Vision Research, 37: 3311-3325.

[34] Pawlak, Z. (1999). Rough sets – theoretical aspects of reasoning about data. Boston, London, Dordrecht: Kluwer Academic Publishers.

[35] Pawlak, Z. (2003). A rough sets view on Bayes' theorem. Int J Intelligent Systems, 18, 487-498.

[36] Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. Nature, 343, 263–266.

[37] Pollen, D. A., Przybyszewski, A. W., Rubin, M. A., & Foote, W. (2002). Spatial receptive field organization of macaque V4 neurons. Cerebral Cortex, 12, 601–616.

[38] Pollen, D. A., Ronner , S. F. (1981) Phase relationships between adjacent simple cells in the visual cortex. Science 212, 1409-1.

[39] Polkowski L, Skowron A. (1996) Rough mereology: A new paradigme for approximate reasoning. Int J Approx. Reasoning, 15 : 333-365.

[40] Przybyszewski, A. W. (1991) An analysis of the oscillatory patterns in the central nervous system with the wavelet method. J Neurosci Methods. 38:247-57.

[41] Przybyszewski AW. (1998) Vision: does top-down processing help us to see? Curr Biol. 8: R135-9.

[42] Przybyszewski, A. W. (2010). Logical rules of visual brain: From anatomy through neurophysiology to cognition. Cognitive System Research 11, 53-66.

[43] Przybyszewski, A. W., Gaska, J. P., Foote, W., & Pollen, D. A. (2000). Striate cortex increases contrast gain of macaque LGN neurons. Visual Neuroscience, 17, 485–494.

[44] Przybyszewski, A. W., Kagan. I., & Snodderly, M. (2003). Eye positioninfluences contrast responses in V1 of alert monkey [Abstract]. Journalof Vision, 3, 698, 698a, <http://journalofvision.org/3/9/698/>.

[45] Przybyszewski, A. W., & Kon, M. A. (2003). Synchronization-based model of the visual system supports recognition. Program No. 718.11. 2003Abstract Viewer/Itinerary Planner. Washington, DC: Society for Neuroscience, Online.

[46] Przybyszewski AW, Lankheet MJ, van de Grind WA. (1993) Nonlinearity and oscillations in X-type ganglion cells of the cat retina.Vision Res. 33: 861-75.

[47] Przybyszewski AW, Lankheet MJ, van de Grind WA. (1996) On the complex dynamics of intracellular ganglion cell light responses in the cat retina. Biol Cybern. 74:299-308.

[48] Przybyszewski, A. W., Linsay, P. S., Gaudiano, P., & Wilson, C. (2007). Basic difference between brain and computer: integration of asynchronous processes implemented as hardware model of the retina. IEEE Transactions of Neural Networks, 18, 70–85.

[49] Przybyszewski, A.W., Potapov, D.O., Rockland, K.S. (2001) Feedback connections from area V4 to LGN. In: Ann. Meet. Society for Neuroscience, San Diego, USA http://sfn.scholarone.com/itin2001/prog#620.9

[50] Ramachandran, V. S. (1988). Perception of shape from shading. Nature, 331, 163–166.

[51] Rao RP, Ballard DH. (1997) Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput. 9(4): 721-63.

[52] Riesenhuber M, Poggio T. (2002) Neural mechanisms of object recognition. Curr Opin Neurobiol. 12:162-8.

[53] Rockland, K. S. (1992). Configuration, in serial reconstruction, of individual axons projecting from area V2 to V4 in the macaque monkey. Cerebral Cortex, 2, 353–374.

[54] Rockland, K. S., Saleem, K. S., & Tanaka, K. (1994). Divergent feedback connections from areas V4 and TEO in the macaque. Visual Neuroscience, 11, 579–600.

[55] Rockland, K. S., & Virga, A. (1990). Organization of individual cortical axons projecting from area V1 (area 17) to V2 (area 18) in the macaque monkey. Visual Neuroscience, 4, 11–28.

[56] Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. Nature Neuroscience, 5, 629–630.

[57] Russell, S., & Norvig, P. (2003). Artificial intelligence: A modern approach (2nd ed.). Prentice Hall Series in Artificial Intelligence.

[58] Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. Journal of Neurophysiology, 39, 1288–1319.

[59] Schummers, J., Marino, J., & Sur, M. (2002). Synaptic integration by V1 neurons depends on location within the orientation map. Neuron, 36, 969–978.

[60] Schummers J, Sharma J, Sur M. (2005) Bottom-up and top-down dynamics in visual cortex. Prog Brain Res. 149:65-81.

[61] Shams, L. & von der Malsburg, C. (2002) The role of complex cells in object recognition. Vision Research 42: 2547-2554.

[62] Sherman, S. M., & Guillery, R. W. (1998). On the actions that one nervecell can have on another: distinguishing ‘‘drivers” from ‘‘modulators”. Proceedings of the National Academy of Sciences of the United States of America, 95, 7121–7126.

[63] Y. Sugase, S. Yamane, S. Ueno, and K. Kawano, (1999). Global and fine information coded by single neurons in the temporal visualcortex. Nature 400, 869–873.

[64] Thorpe, S., Faze, D., & Merlot, C. (1996). Speed of processing in the human visual system. Nature, 381, 520–522.

[65] Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. Quarterly Journal of Experimental Psychology A, 40, 201–237.

[66] Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nature Neuroscience, 5, 682–687.

[67] Wang D, Terman D. Image segmentation based on oscillatory correlation.Neural Comput. 1997 9: 805-36.

[68] Zadah, L. A., "Toward a perception-based theory of probabilistic reasoning with imprecise probabilities", Journal of Statistical Planning and Inference, 2002,105, pp. 233-264.

[69] Zadah, L. A., From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions, International Journal of Applied Math and Computer Science, vol. 12, no. 3, pp. 307-324, 2002.